

# A New Look at Data Usage by Using Metadata Attributes as Indicators of Data Quality

Young-In Won<sup>1,3</sup> ([youngin.won@nasa.gov](mailto:youngin.won@nasa.gov)), Lalit Wanchoo<sup>2,3</sup> ([lalit.wanchoo@nasa.gov](mailto:lalit.wanchoo@nasa.gov)), Jeanne Behnke<sup>3</sup> ([jeanne.behnke-1@nasa.gov](mailto:jeanne.behnke-1@nasa.gov))

<sup>1</sup> KBRwyle, 7515 Mission Drive, Suite A100, Lanham, MD 20706, <sup>2</sup> ADNET Systems, Inc., 7515 Mission Drive, Lanham, MD 20706, <sup>3</sup> Earth Science Data and Information System Project (Code 423), NASA Goddard Space Flight Center, Greenbelt, MD 20771

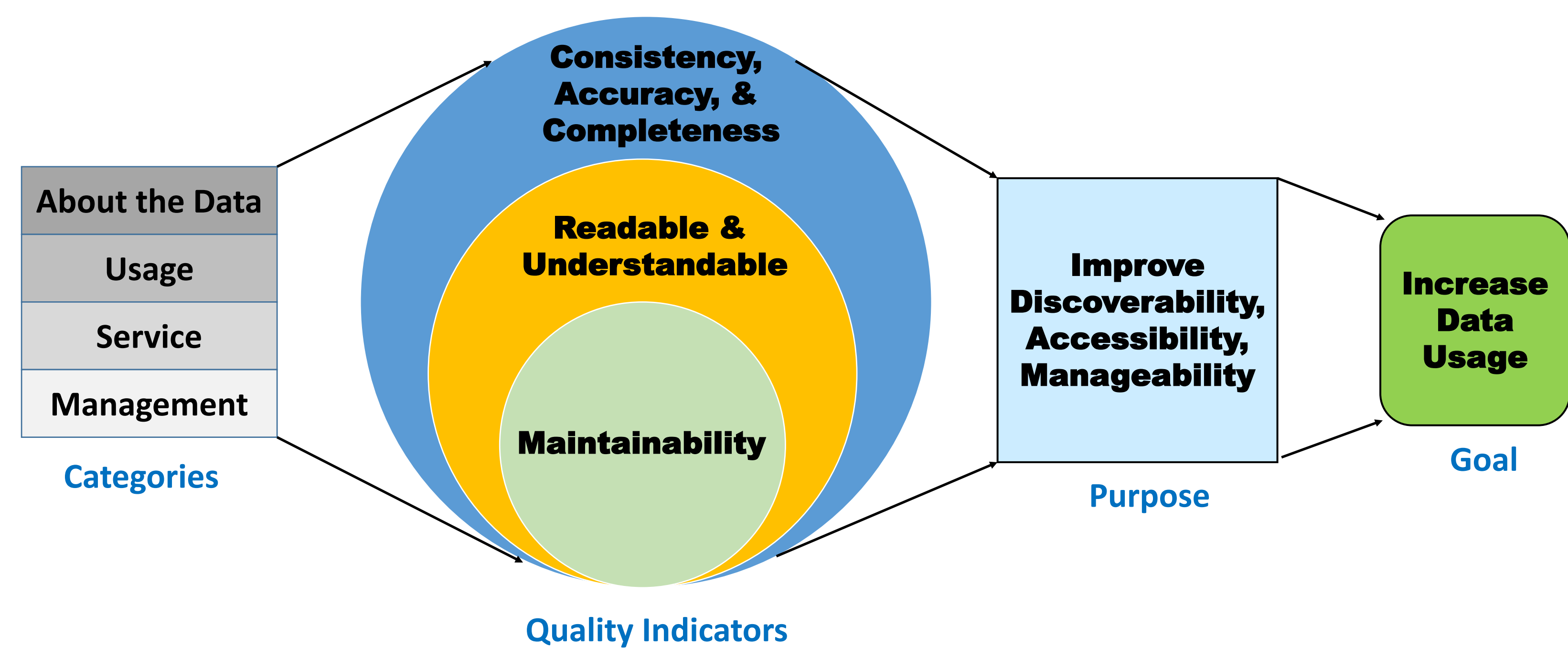
## Introduction

NASA's Earth Observing System Data and Information System (EOSDIS) stores and distributes data from EOS satellites, as well as ancillary, airborne, in-situ, and socio-economic data. Twelve EOSDIS data centers support different scientific disciplines by providing products and services tailored to specific science communities. Although discipline-oriented, these data centers provide common data management functions of ingest, archive, and distribution, as well as documentation of their data and services on their web-sites. The Earth Science Data and Information System (ESDIS) Project collects these metrics from the EOSDIS data centers on a daily basis through a tool called the ESDIS Metrics System (EMS). These metrics were used in this study.

The implementation of the Earthdata Login – formerly known as the User Registration System (URS) – across the various NASA data centers provides the EMS additional information about users obtaining data products from EOSDIS data centers. These additional user attributes collected via the Earthdata Login, such as the user's primary area of study, can augment the understanding of data usage, which in turn can help the EOSDIS Project team better understand the users' needs.

This study reviews the key metrics (users, distributed volume, and files) in multiple ways to gain an understanding of the significance of the metadata. Characterizing the usability of data by key metadata elements, such as discipline and study area, will assist in understanding how the user's needs have evolved over time. The data usage pattern based on product level provides insight into the level of data quality. In addition, the data metrics by various services, such as the Open-source Project for a Network Data Access Protocol (OPeNDAP) and subsets, address how these services have extended the usage of data. Over-all, this study presents the usage of data and metadata by metrics analyses, which may assist data centers in better supporting the needs of the users.

## Metadata Workflow

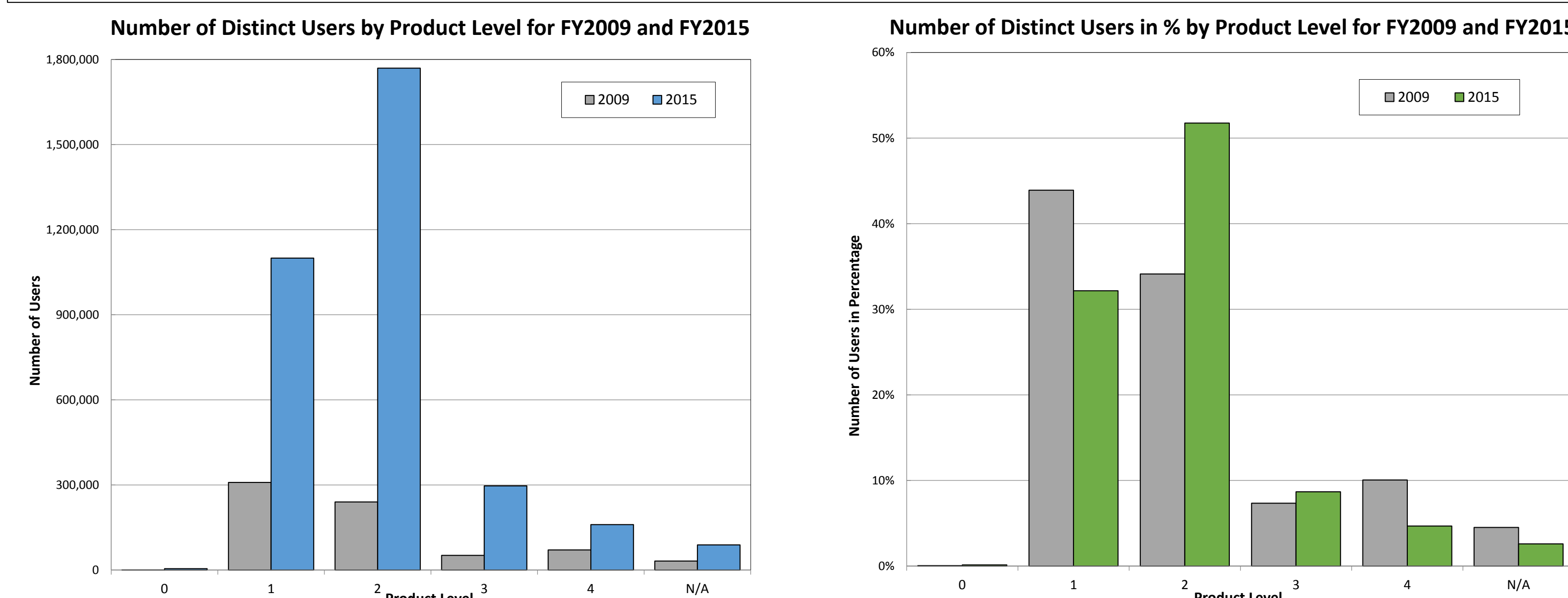


- Providing as much detailed metadata of improved quality should enhance discoverability, accessibility, and manageability.
- Easier for users to find the data → create higher demands for the data.

## Characterizing Usability of Data by Key Metadata

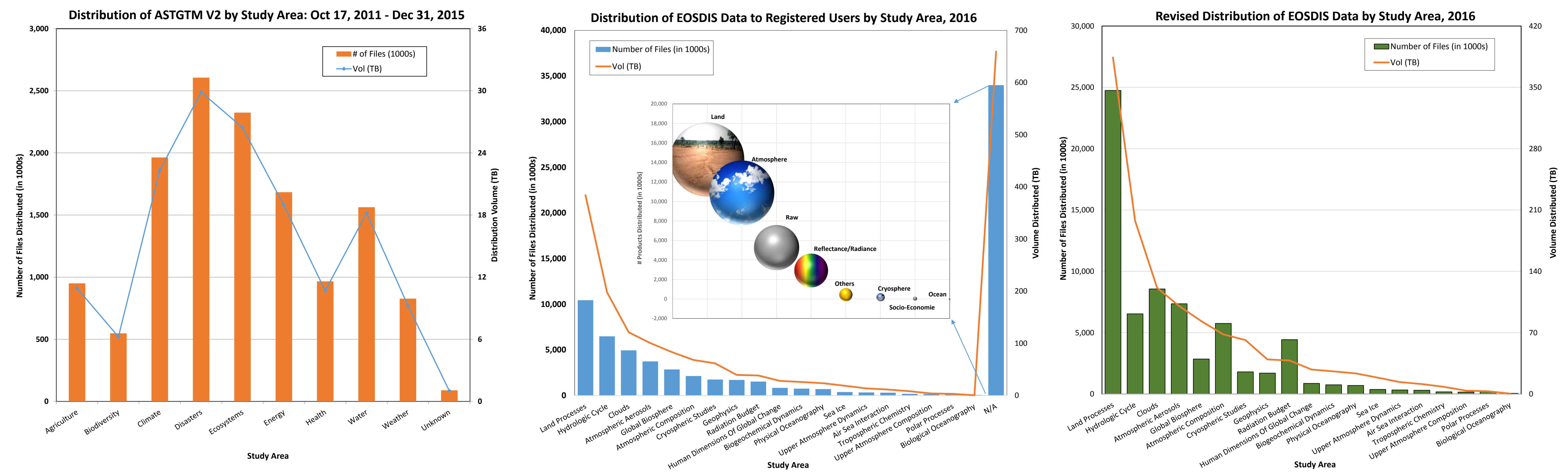
- The data usage and the number of users accessing data can be analyzed in various ways by using metadata attributes such as (some of these are specific to EMS): Mission, Instrument, Product Level, File Version, Latitude/Longitude, Country, User Type, Study Area (Science Key words/Discipline), Service and Protocol, etc.

## Metadata Attribute : Product Level



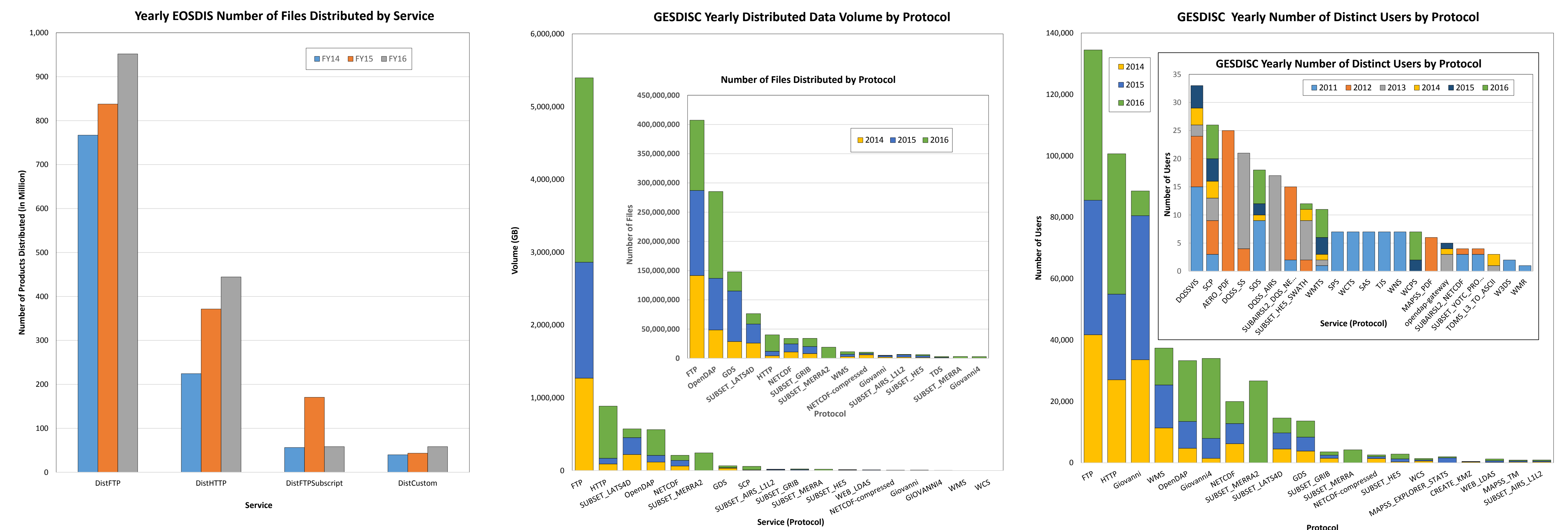
- Number of distinct users (based on IPhost) by product level for all EOSDIS data distributed during two different years (left: absolute number, right: in percentage)
- Improvement of metadata, better identification of products → easier for user to find the data
- Most downloaded data products shifting from Level 1 to Level 2, more usage of derived geophysical quantities than lower level radiance → may indicate the improved data quality with time.

## Metadata Attribute : Study Area, Discipline



- NASA Land Processes Distributed Active Archive Center (LP DAAC) started adding a metadata “Study Area” in its distribution log for Aster Global Digital Elevation Model data to EMS from the beginning of its data release (left).
- Metadata “Study Area” was added to the ESDIS User Registration System (Earthdata Login), thereby improving users characterization.
- With the improved user characterization, the distribution metrics based on “Study Area” can now be applied to the whole EOSDIS data.
- There are still users whose study area is not known (“N/A” in the center figure). For such users, “Science Discipline” may be substituted.

## Metadata Attribute : Service and Protocol



- Number of distributed data products by service (the traditional way, left)
- The addition of a new field called “Protocol” from Goddard Earth Sciences Data and Information Services Center (GESDISC) in the distribution log → enabling detailed data distribution metrics by protocol (center and right)
- Usage of various protocols in the data distribution — with time depicting “OPeNDAP”, “HTTP”, and some subset services — became more popular with time.
- The most and least popular services depicted (right) → should be of great help for the data providers in deciding where to focus their efforts in maintaining various services, and/or deciding which service to terminate.

## Summary

- ❖ The data discoverability, accessibility, and manageability has been enhanced by improving the key metadata as shown in the increase in usage.
- ❖ The EMS has evolved with time through the addition of a number of new metadata attributes, which has enabled value-added metrics analysis on the usage of data.
- ❖ The addition of new metadata elements, such as “Study Area”, assists in analyzing the usage of data products.
- ❖ The metrics based on service (protocol) clearly demonstrate how each service has extended over time and may help data providers in deciding where to focus their efforts to better serve users' needs.

## Acknowledgments

The contributions of Young-In Won, and Lalit Wanchoo to this study were funded through the Science and Exploration Data Analysis (SESDA III) GSFC NASA Contract No: NNG12PL17C, and the contributions of the Deputy Project Manager/Operations of the ESDIS Project Jeanne Behnke' were provided by NASA.